Housing Predictions Executive Summary

Our Housing Data project focuses on using NTREIS[1](North Texas Real Estate Information Service) data and machine learning to predict the closing price of potential sellers' houses. We also used Census Data[2] and Tableau[3] to help visualize the demographics of the areas that we are analyzing with our housing data.

By using the NTREIS data, we were able to run several machine learning models to accurately predict housing prices. We reviewed the data to identify the columns and rows we wanted to keep and discard. Once our data was clean, we then had to process our data to make our prediction models run efficiently. This is done by one-hot encoding, label encoding, and scaling our data as needed.

Once processed, we reviewed our correlations to see which features would help us make accurate predictions. The features we decided on were: county, city, HOA type, square feet, bedroom total, and bathroom total. These features had the highest correlation which means these will be the most helpful in accurately predicting housing prices. From there, we were able to run our models to determine which model predicts the most accurately. After running several models, we chose gradient boost as our official model that we want to run our test data on and make our predictions. We also attempted to predict days on market, but with every model we tried, we received negative r-squared values, which indicates a very poor prediction. After everything we tried, we were not able to resolve this issue, therefore dropped this prediction.

Tableau was utilized to visualize all of the data we worked with as well as the Census data information. We made multiple dashboards allowing the user to interactively see historical housing transactions which will allow the user to have a better idea what buys are looking for. The Census dashboard also gives users an idea of the demographics of each DFW area and how that may impact the sale price of a house.

We created a web application that allows the user to input their house information, press a button, and receive a real-time prediction on the selling price for their house using the model we selected (gradient boosting). The user can then go a step further and explore several visualizations of the DFW area housing data and demographic information as well. Each page of our web application allows the user to interact and discover different facts and features of the DFW market. The added instructions make for a user-friendly experience. We used red, white, and blue as our theme as many real estate sites (NTREIS included) use these colors.

From our analysis we were able to determine that the more features a house has (e.g., rooms, bathrooms, garages, pools, etc.) the more expensive the house is. The location of the house has a definite impact on the price of the house as well.

Due to the limited data in our dataset, we weren't able to analyze housing data in other areas of Texas or the US, so in the future we would like to expand our dataset to incorporate other areas.

1 - https://ntreis.net/accessing-ntreis-data/accessing-ntreis-data-agents/

2 - Texas - Census Bureau Tables

3 - https://public.tableau.com/app/profile/us.census.bureau/viz/2020CensusPopulationandHousingMap/AllStates

Brandilyn Hall, Elyssa Irizarry, Megan Adams, Yvonne Martinez
Data Science Bootcamp
Project 4 - Group 1

This will also allow us to have a more robust dataset instead of just the 5000 rows over the last 6 months.

This web application can give users useful information on how to prepare their house for sale, what to invest money in, what to expect, and even when to list the house.